<u>REMARKS</u>

This application has been reviewed in light of the Office Action dated January 29, 2004. Claims 1, 3-16, 18, 20-33, 35, 37-50, 52, 54-66, 68-80, and 82-96 are presented for examination, of which Claims 1, 18, 35, 52, 66, 80, and 94-96 are in independent form. Claims 17, 34, and 51 have been canceled, without prejudice or disclaimer of matter, and will not be discussed further. Claims 1, 9, 18, 26, 35, 43, 52, 63-66, 77-80, and 91-94 have been amended to define still more clearly what Applicant regards as her invention. Claims 95 and 96 have been added to provide Applicant with a more complete scope of protection. Favorable reconsideration is requested.

Claims 1, 3-10, 13-16, 18, 20-27, 30-33, 35, 37-44, 47-50, 52, 54-61, 66, 68-75, 80, 82-89, and 94 were rejected under 35 U.S.C. § 103(a) as being unpatentable over IEEE Paper ISBN: 0162-8828; "A Markov Random Field Model-Based Approach to Image Interpretation (*Modestino et al.*) in view of U.S. Patent No. 6,360,234 (*Jain et al.*), and Claims 11, 28, 45, 62-65, 76-79, and 90-93 were rejected under Section 103(a) as being unpatentable over *Modestino et al.* in view of *Jain et al.* and U.S. Patent No. 5,930,783 (*Li et al.*).

As shown above, Applicant has amended independent Claims 1, 18, 35, 52, 66, 80, and 94 in terms that more clearly define what she regards as her invention. Applicant submits that these amended independent claims and new independent Claims 95 and 96, together with the remaining claims dependent thereon, are patentably distinct from the cited prior art for at least the following reasons.

The aspect of the present invention set forth in Claim 1 is a method of classifying a digital image. The method segments the digital image into a plurality of substantially homogeneous regions and processes the plurality of regions to provide a

region adjacency graph for the digital image. The region adjacency graph represents spatial adjacency between the plurality of regions of the digital image. The method also includes labeling at least one of the regions of the region adjacency graph with one of a plurality of predetermined semantic labels to provide a labeled region adjacency graph. The labeled region adjacency graph is analyzed to identify one or more predetermined patterns of the semantic labels in the labeled region adjacency graph. One of a plurality of predetermined stereotypes is assigned to the labeled region adjacency graph according to at least one identified predetermined pattern of the semantic labels in the labeled region adjacency graph. Each of the predetermined stereotypes corresponds to at least one of the predetermined patterns, such that the assigned stereotype describes the plurality of regions of the digital image and represents a classification of the digital image. The method further includes storing the assigned stereotype and the digital image in one or more databases of digital images, wherein the digital image is retrievable from the one or more databases using the assigned stereotype.

Among other important features of Claim 1 is assigning one of a plurality of predetermined stereotypes to the labeled region adjacency graph according to at least one identified predetermined pattern of the semantic labels in the labeled region adjacency graph, each of the predetermined stereotypes corresponds to at least one of the predetermined patterns, such that the assigned stereotype describes the plurality of regions of the digital image and represents a classification of the digital image.

Another important feature of Claim 1 is storing the assigned stereotype and the digital image in one or more databases of digital images, wherein the digital image is retrievable from the one or more databases using the assigned stereotype.

As described at page 7, lines 1-7, of the specification[1], at each analysis event, the video frame centered on the analysis event is automatically spatially segmented into homogeneous regions. These regions and their spatial adjacency properties are represented by a Region Adjacency Graph (RAG). The probabilistic model is then applied to the RAG. The model incorporates feature measurements from the regions of the frame, contextual information from a Region of Interest (ROI) around the frame, and prior knowledge about the various semantic labels that can be associated with the regions of the RAG. Further, ass described at page 7, lines 12-15, higher-level expressions termed stereotypes can be used to represent classifications of the video frame or image. Stereotypes can be assigned to video frames, or images, by detecting patterns of region labels and (their) corresponding adjacency in the RAG. This assignment requires a predetermined set of patterns of region labels to be provided, wherein each pattern is associated with a stereotype.

As described at page 7, lines 7-12, the probabilistic model incorporates feature measurements from the regions of the frame, contextual information from a Region of Interest (ROI) around the frame, and prior knowledge about the various semantic labels that can be associated with the regions of the RAG. These semantic labels (eg., "person", "sky", "water", "foliage", etc.) are taken from a list which has been typically constructed for an appropriate application domain (e.g., outdoor scenes, weddings, urban scenes, etc).

In the Response to Arguments section of the Office Action and in making the current rejection under 35 U.S.C. 103, the Examiner states that the joint assignment of

---

[1]It is to be understood, of course, that the claim scope is not limited by the details of the described embodiments, which are referred to only to facilitate explanation.

object labels for all the image primitives provides a stereotype for interpretation of the image.

However, Applicant submits that nothing has been found in *Modestino et al.* that would teach or suggest assigning one of a plurality of predetermined stereotypes to the labeled region adjacency graph according to at least one identified predetermined pattern of the semantic labels in the labeled region adjacency graph, such that the assigned stereotype describes the plurality of regions of the digital image (i.e., the entire region adjacency graph and the entire digital image) and represents a classification of the digital image. As discussed above, stereotypes can be used to represent classifications of the video frame or image. Stereotypes can be assigned to video frames, or images, by detecting patterns of region labels and (their) corresponding adjacency in the RAG. This assignment requires a predetermined set of patterns of region labels to be provided, wherein each pattern is associated with a stereotype (page 7, lines 12-15, of the specification).

In contrast, in *Modestino et al.* interpretation labels denoted by *I* are assigned to the segmented regions (i.e., one label to one region) using domain knowledge, extracted feature measurements, and spatial relationships between the various regions (Abstract, lines 3-10). At lines 1-3 of Section II-B of *Modestino et al.,* the image interpretation problem described in Section II-B is restricted to that of labeling the segmented regions. *Modestino et al.* then discusses at lines 9-11 of Section II-B that a neighborhood system *n* and, consequently, a set of cliques can also be defined on the adjacency graph. In Fig. 1, and as described at pages 607 and 608, Section II-B, *Modestino et al.* shows the adjacency graph and all its cliques for a particular synthetic image, as one example. *Modestino et al.* describes in Section II-B the determination of the vector *I(R)*

representing the object labels assigned to the segmented regions and the optimization of the vector $I(R)$ represented by $Io(R)$.

The Examiner appears to equate the perceived two regions and three regions categories depicted in Table 1 in association with Figs. 1(a) to 1(c) of *Modestino et al.* with stereotypes as defined in the present application at page 7, lines 11-17, (i.e., a higher-level expression representing a classification of the video frame or image) and as claimed in independent Claim 1. The two and three region categories depicted in Table 1 of *Modestino et al.* represent domain knowledge, as disclosed at page 611, column 1, lines 4-7. Domain knowledge is applied by *Modestino et al.* during the process of labeling the region adjacency graph, as described with reference to Equation 4, at Section II-B, pages 607 and 608.

However, the particular claimed limitation of the present invention of assigning one of a plurality of predetermined stereotypes to the labeled region adjacency graph according to at least one identified predetermined pattern of the semantic labels in the labeled region adjacency graph, each of the predetermined stereotypes corresponding to at least one of the predetermined patterns such that the assigned stereotype describes the plurality of regions of the digital image and represents a classification of the digital image, greatly improves search and retrieval of digital images (page 15, line 12, to page 16, line 6, of the present specification).

The method of interpretation of *Modestino et al.* and the interpretation of the term stereotype given by the Examiner would not produce the same results as the present invention. For example, an image may classified as a beach scene, in accordance with the present invention, based on its labeled RAG containing labels sky, water and a sand in a certain pattern (see Table 1 of the present specification). If a further labeled

region (e.g., a road) was present and/or added to the RAG (e.g., adjacent to the sand) for the image, then the image would still be assigned the same classification (i.e., a beach scene), since the same pattern of labels still exists in the RAG. In contrast, using the method of *Modestino et al.*, based on the Examiner's interpretation of the word stereotype, the presence of a further labeled region to a RAG would produce a different classification of the image associated with the RAG, since in the Examiner's interpretation it is the joint assignment of object labels for all the image primitives that provides a stereotype for interpretation of the image.

Applicant submits that the combination of *Modestino et al.* and *Jain et al.*, assuming such a combination would be permissible, still would not teach or suggest the feature of Claim 1 of assigning one of a plurality of predetermined stereotypes to the labeled region adjacency graph according to at least one identified predetermined pattern of the semantic labels in the labeled region adjacency graph, each of the predetermined stereotypes corresponding to at least one of the predetermined patterns such that the assigned stereotype describes the plurality of regions of the digital image and represents a classification of the digital image. Applicant submits that a combination of *Jain et al.* and *Modestino et al.* would produce a classification and search system where interpretation labels would be assigned to segmented regions of an image using domain knowledge, extracted feature measurements and spatial relationships between the various regions. These labels may be used in a semantic based query in order to retrieve the image. However, in accordance with the Examiner's interpretation, either all of the labels (i.e., the joint of assignment of the labels representing a classification of the image) would be required to be included in the query to retrieve the image, or each of the labels individually would result in the image being retrieved.

In making the rejection of Claim 17 under 35 U.S.C. 103(a), the Examiner states that *Modestino et al.* and *Jain et al.* do not explicitly disclose that the digital image is stored in a database of digital images wherein the classification can be used to retrieve the digital image from the database. Applicant agrees. However, the Examiner contends that *Li et al.* discloses a semantic and cognition based image retrieval methodology wherein the digital image is stored in a database of digital images and wherein the classification can be used to retrieve said digital image from the database (column 4, lines 32-50).

At column 4, lines 32-50, *Li et al.* discusses three types of queries (i.e., semantics based, cognation specification and object spatial relationship specification) for selecting images and a four step method for processing those queries. However, nothing has been found in *Li et al.* that would teach or suggest storing the assigned stereotype and a digital image in one or more databases of digital images, where the digital image is retrievable from the one or more databases using the one or more assigned stereotypes.

Accordingly, Applicant submits that neither *Modestino et al.*, *Jaine et al.*, *Li et al.* nor any combination thereof (assuming *arguendo* that any such combination would be permissible) would teach or suggest the particular claimed feature of Claim 1 of assigning one of a plurality of predetermined stereotypes to the labeled region adjacency graph according to at least one identified predetermined pattern of the semantic labels in the labeled region adjacency graph, each of the predetermined stereotypes corresponding to at least one of the predetermined patterns such that the assigned stereotype describes the plurality of regions of the digital image and represents a classification of the digital image. Further, Applicant submits that neither *Modestino et al.*, *Jaine et al.*, *Li et al.* nor any combination thereof would teach or suggest and storing the assigned stereotypes and the digital image in one or more databases of digital images, wherein the digital image is

retrievable from the one or more databases using the assigned stereotype, as recited in Claim 1.

A combination of *Modestino et al.*, *Jaine et al.*, and *Li et al.* would produce a classification, search and retrieval system where interpretation labels would be assigned to segmented regions of an image using domain knowledge, extracted feature measurements and spatial relationships between the various regions. These labels may be used in a semantic based query in order to retrieve the image. However, as discussed above, in accordance with the Examiner's interpretation, either all of the labels (i.e., the joint of assignment of the labels representing a classification of the image) would be required to be included in the query to retrieve the image, or each of the labels individually would result in the image being retrieved. In either instance, the image retrieval system would be highly inefficient. For example, in the second instance, for a database of images classified in such a manner, a first label would result in many different images being retrieved. A different label would again result in many different images being retrieved with some if not all of the images being the same as those retrieved for the first label even though a different label was used.

Accordingly, Applicant submits that Claim 1 is clearly patentable over the cited prior art.

Independent Claims 18 and 35 are apparatus and computer program product claims, respectively, corresponding to method Claim 1, and are believed to be patentable for at least the same reasons as discussed above in connection with Claim 1. Additionally, independent Claims 52, 66, and 80 include similar features as discussed above in connection with Claim1, and are also believed to be patentable for reasons substantially similar as those discussed above in connection with Claim 1.

The aspect of the present invention set forth in Claim 94 is a method of classifying a digital image. The method comprises the steps of segmenting the image into substantially homogeneous regions. The regions are processed to provide a labeled region adjacency graph for the digital image. The labeled region adjacency graph represents spatial adjacency between the regions of the image, at least one of the regions of the labeled region adjacency graph being associated with one of a plurality of predetermined semantic labels. The labeled region adjacency graph is analyzed to identify one or more predetermined patterns of the semantic labels in the labeled region adjacency graph. One or more of a plurality of predetermined stereotypes are assigned to the digital image according to each identified predetermined pattern of semantic labels in the labeled region adjacency graph. The plurality of predetermined stereotypes are represented by a multi-level hierarchical structure with each of the predetermined stereotypes of the hierarchical structure corresponding to at least one of the predetermined patterns such that the one or more assigned stereotypes represent a classification of the digital image. The one or more assigned stereotypes and the digital image are stored in one or more databases of digital images together with one or more hierarchical paths. The one or more hierarchical paths are based on the multi-level hierarchical structure. The digital image is retrievable from the one or more databases using the one or more assigned stereotypes and hierarchical paths.

Among other important features of Claim 94 are assigning one or more of a plurality of predetermined stereotypes to the digital image according to each identified predetermined pattern of the semantic labels in the labeled region adjacency graph, the plurality of predetermined stereotypes being represented by a multi-level hierarchical structure with each of the predetermined stereotypes of the hierarchical structure

- 35 -

corresponding to at least one of the predetermined patterns such that the one or more assigned stereotypes represent a classification of the digital image, and storing the one or more assigned stereotypes and the digital image in one or more databases of digital images together with one or more hierarchical paths, the one or more hierarchical paths being based on the multi-level hierarchical structure, wherein the digital image is retrievable from the one or more databases using the one or more assigned stereotypes and hierarchical paths

As described at page 15, lines 5-11, the classification of the stereotypes is represented in a hierarchical system so that the search/retrieval at any level of the hierarchy is possible. An example of such a hierarchy for outdoor scenes is shown in Fig. 5. As described at page 15, lines 9 and 10, a hierarchical path can be stored with the stereotype in a metadata object which is associated with the image/video object or accessed by reference. Further, as described at page 15, line 12, to page 16, line 6, the generated stereotypes can form the basis of an image retrieval system. The available stereotypes can be presented to the user in the form of icons. For example, a "portrait" icon shows just a single face taking up the area of the icon, whereas a "crowd" icon represents the possibility of many people/faces. The icons can be presented in the form of icon trees through which the user can navigate, where the icon trees represent the hierarchical arrangement of stereotypes. The user can select the icon representing the stereotype of the desired image(s), and a query can be generated for the request. The query is processed and all images matching the query are retrieved and presented to the user. Such an image retrieval system can be implemented without the costly requirement of manual annotation, since the stereotypes of the present invention are automatically generated by a digital image or digital video interpretation system.

- 36 -

As discussed above, at column 4, lines 32-50, *Li et al.* discusses three types of queries (i.e., semantics based, cognation specification and object spatial relationship specification) for selecting images and a four step method for processing those queries. However, nothing has been found in *Li et al.* that would teach or suggest storing one or more assigned stereotypes and a digital image in one or more databases of digital images together with one or more hierarchical paths, where the one or more hierarchical paths are based on a multi-level hierarchical structure of predetermined stereotypes, as recited in Claim 94. Further, nothing has been found in *Lie et al.* that would teach or suggest that the digital image is retrievable from the one or more databases using the one or more assigned stereotypes and hierarchical paths, as further recited in Claim 94.

In making the rejection under Section 103(a), the Examiner states that *Jain et al.* discusses that each of the stereotypes has a hierarchical arrangement (Figures 7, 9 and 1, column 6, lines 29-47, column 13, lines 52-67, and column 14, lines 1-31). The Examiner also states that *Jain et al.* discusses that each of the stereotypes has a hierarchical path arrangement (Figures 6, 7, 9 and 17, column 6, lines 29-47, column 13, lines 52-67, and column 14, lines 1-31).

Fig. 6 of *Jain et al.* is a logical illustration of a number of metadata types in the form of a time-based track representation. Fig. 6 shows eight tracks, namely, a keyframe track 320, a CC-text track 322, an audio class track 324, a speech track 326, speaker ID track 328, a keyword track 330 a custom track 331 and a clip track 332. Column 6, lines 29-47, of *Jain et al.* describes Fig. 6, in that each of the tracks of Fig. 6 are a parcel of different types of metadata. For example, the keyframe track consists of an individual set of keyframes. Fig. 7 of *Jain et al.* shows an object model for these metadata types with a software process that manages the metadata. Fig. 7 shows the main object, the

- 37 -

Metadata Track Index Manager 402 and different metadata track objects such as keyframes 406, CC-text 408 audio classes 410, speech 412 and video session level metadata 404 holding the different types of metadata. As disclosed at column 7, lines 10-14, the video session level metadata 404 is where the information for managing and time-synching the encoded video message resides. However, the metadata types of *Jain et al.* are not represented in Figs. 6 and 7 of *Jain et al.* in a multi-level hierarchical.

Fig. 9 of *Jain et al.* shows the main architectural elements of an extensible Video Engine 440 disclosed in *Jain et al.* Incoming media is processed by a Media Capture Services 430 consisting of Timecode Capture 502, Video Capture 504, Audio Capture 506 and Text Capture 508. Digital media 509 is then made available to Feature Extractor Framework 510 for processing. However, there is no teaching or suggestion in Fig. 9 of *Jain et al.* of stereotypes represented by a multi-level hierarchical structure.

Fig. 16 of *Jain et al.* is a flowchart of a HTML output filter process. Fig. 17 of *Jain et al.* is an exemplary screen display seen as an output of the output filter process of Fig. 16. Column 13, lines 52-67, and column 14, lines 1-31, of *Jain et al.* describe Figs. 16 and 17. Fig. 17 shows a video frame 896, keyframe 904, a CC-text frame 908 and a clip frame 912 within a browser window 916.

Applicant submits that the above cited figures and passages of *Jain et al.* fail to teach or suggest the features of Claim 94.

Applicant submits that neither *Jain et al.*, *Modestino et al.*, *Li et al.* nor any combination thereof would teach or suggest the particular claimed features of Claim 94 of assigning one or more of a plurality of predetermined stereotypes to the digital image according to each identified predetermined pattern of the semantic labels in the labeled region adjacency graph, the plurality of predetermined stereotypes being represented by a

- 38 -

multi-level hierarchical structure with each of the predetermined stereotypes of the hierarchical structure corresponding to at least one of the predetermined patterns such that the one or more assigned stereotypes represent a classification of the digital image. Applicant further submits that neither *Jain et al.*, *Modestino et al.*, *Li et al.* nor any combination thereof would teach or suggest storing the one or more assigned stereotypes and the digital image in one or more databases of digital images together with one or more hierarchical paths, the one or more hierarchical paths being based on the multi-level hierarchical structure, wherein the digital image is retrievable from the one or more databases using the one or more assigned stereotypes and hierarchical paths, as further recited in Claim 94.

Accordingly, Applicant submits that Claim 94 is clearly patentable over the cited prior art.

Applicant submits that neither *Jain et al.*, *Modestino et al.*, *Li et al.* nor any combination thereof would teach or suggest the particular claimed features of Claim 95 of the plurality of stereotypes being represented in a multi-level hierarchical tree-structure such that the digital image is retrievable from one or more databases of digital images upon selection of at least one of the assigned stereotypes using the multi-level hierarchical tree-structure. For example, as described at page 15, line 12, to page 16, line 6, the generated stereotypes can form the basis of an image retrieval system. The available stereotypes can be presented to the user in the form of icons. A "portrait" icon shows just a single face taking up the area of the icon, whereas a "crowd" icon represents the possibility of many people/faces. The icons can be presented in the form of icon trees through which the user can navigate, where the icon trees represent the hierarchical arrangement of stereotypes. The user can select the icon representing the stereotype of the desired

image(s), and a query can be generated for the request. The query is processed and all images matching the query are retrieved and presented to the user.

Accordingly, Applicant submits that Claim 95 is clearly patentable over *Jain et al.*, *Modestino et al.*, *Li et al.*, taken separately or in any permissible combination.

Independent Claim 96 includes features similar to those discussed above in connection with Claim 95, and is also believed to be patentable for reasons substantially similar as those discussed above in connection with Claim 95.

A review of the other art of record has failed to reveal anything which, in the Applicant's opinion, would remedy the deficiency of the art discussed above, as references against the independent claims herein. Those claims are therefore believed to be patentable over the other art of record.

The other claims in this application are each dependent from one or another of the independent claims discussed above and are therefore believed patentable for the same reasons. Since each dependent claim is also deemed to define an additional aspect of the invention, however, the individual reconsideration of the patentability of each on its own merits is respectfully requested.

This Amendment After Final Action is believed clearly to place this application in condition for allowance and, therefore, its entry is believed proper under 37 C.F.R. § 1.116. Accordingly, entry of this Amendment After Final Action, as an earnest effort to advance prosecution and reduce the number of issues, is respectfully requested. Should the Examiner believe that issues remain outstanding, it is respectfully requested that the Examiner contact Applicant's undersigned attorney in an effort to resolve such issues and advance the case to issue.

In view of the foregoing amendments and remarks, Applicant respectfully requests favorable reconsideration and early passage to issue of the present application.

Applicant's undersigned attorney may be reached in our New York office by telephone at (212) 218-2100. All correspondence should continue to be directed to our below listed address.

Respectfully submitted,

Attorney for Applicant

Registration No. 29,296

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200

NYMAIN431368